1989, Gene Anal Techn 6:25–28

Use of Avian Myeloblastosis Virus Reverse Transcriptase at High Temperature for Sequence Analysis of Highly Structured RNA

ELAINE SHIMOMAYE and
MARIA SALVATO

Avian myeloblastosis virus reverse transcriptase (AMV RT) is routinely used in the sequence analysis of RNA and DNA templates. We review the various methods for dealing with secondary structures that would otherwise result in premature termination or sequence compression. Based on our experience in sequencing the 11-kb single-stranded RNA genome of lymphocytic choriomeningitis virus, we have found that raising the reaction temperature above 47°C is the simplest way to overcome template secondary structure, and the use of 98% formamide gels is the simplest way to overcome product secondary structure.

We have used avian myeloblastosis virus reverse transcriptase (AMV RT) for direct RNA sequence analysis of a single-stranded RNA virus, lymphocytic choriomeningitis virus (LCMV). We wanted an RNA sequencing method that would allow us to sequence several viral variants in tandem and that would circumvent the need to clone viral cDNA.

In performing dideoxy sequence analysis with AMV RT we encountered two types of secondary structure problems: those involving template structures that terminate transcription, and those involving product (cDNA) structures that compress C or G tracts of sequence. Several approaches for overcoming template secondary structure have been described in the literature: treatment of RNA templates with 20 mM methyl mercury [1, 2], use of dimethyl sulfoxide in reactions with double-stranded reovirus RNA templates [3], addition of T4 gene 32 protein (a single-stranded binding protein) to sequencing reactions [4], and use of high reaction temperature (50°C) to sequence tRNA genes in chromosomal DNA [5]. Sequence compressions in the reaction products have been overcome by electrophoresis on gels containing 98% formaldehyde [6] or by incorporating nucleotide analogues that are not prone to intrachain hydrogen bonding [7, 8].

Chemical methods of sequence analysis are less susceptible to template secondary structure than the enzymatic methods. Auperin et al. [9] end-labeled LCMV RNA and performed base-specific cleavage reactions [10] in order to sequence the termini of the virus. We have also used chemical sequencing [11] on difficult regions of the LCMV genome, i.e., regions in which template secondary structure caused premature terminations with AMV RT. This involved AMV RT extension of an end-labeled primer, followed by purification of the longest cDNA products and chemical cleavage of these products to obtain sequence [11]. Unfortunately, the template purification steps required for the chemical sequencing methods makes them much more difficult than the enzymatic chain terminating methods with labeled primers. The enzymatic method is ideal for impure templates such as whole chromosomal DNA or whole cellular RNA because a labeled primer will anneal to and extend a specific sequence. To obtain RNA sequence directly from either viral genomic RNA or infected cellular RNA we chose to use enzymatic chain termination with additional denaturants in the event of problematic secondary structure.

Methods

Oligonucleotide Primers

Our reactions made use of radioactive oligonucleotides, generally 20 nucleotides in length, synthesized on an Applied Biosystems model 380A DNA Synthesizer. Oligonucleotides were deblocked in ammonia for 10 hours at 50°C, dried in a spin vacuum, and resuspended in 0.5 ml distilled water and butanol extracted four times until volume was 50–100 μl. They were dried once again, resuspended in water, and 5 μl (approximately 10 A_{260}U) was loaded onto a thin 20% acrylamide gel for preparative electrophoresis [12]. The oligonucleotide was visualized by transferring the gel on cellophane to a fluorescent silica gel plate (60.F-254, EM Science). The absorbent band was excised with a razor blade and placed under 1 ml TE (10 mM Tris-HCl, pH 8, 1

From the Department of Immunology, Scripps Clinic and Research Foundation, La Jolla, California.

Address reprint requests to: Dr. Maria Salvato, Department of Immunology, Scripps Clinic and Research Foundation, 10666 N. Torrey Pines Road, La Jolla, CA 92037.

Received June 28, 1988.

0735-0651/89/$03.50

1989, Gene Anal Techn 6:25–28

mM EDTA) at 4°C to elute overnight. The next day the gel slice was rinsed with an additional 0.5 ml TE, and the eluate was passed through a disposable 0.45 μm nitrocellulose filter to remove acrylamide pieces. The eluted oligonucleotide was made 0.1 M Na-acetate, 15 μg/ml in glycogen (carrier) and precipitated with three volumes of ethanol. The pellet was suspended in TE, and the oligonucleotide concentration was determined using an extinction coefficient of 20 (i.e., dilution × A260 × 20 = μg/ml). The molar conversion of 7000 g/mole for a 20 mer or 140 pmoles/μg was used. A 20 μM oligonucleotide stock was made and kinased as described [11]. A typical kinase reaction contained 50 pmoles oligonucleotide, 5 units kinase, 50 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM dithiothreitol, 0.1 mM spermidine, 0.1 mM EDTA, 10 μCi γ-^{32}P-ATP, and was incubated for 30 minutes at 37°C. This reaction was phenol extracted, then phenol:chloroform (1:1) extracted, and then ethanol precipitated in the presence of 5 μg glycogen. The pellet was resuspended at 2×10^5 cpm/μl, and 2.5 μl was used per sequencing reaction.

AMV RT Reaction Conditions

The RNA template was denatured with the radioactive oligonucleotide at 100°C for 3 minutes and allowed to anneal for 30 minutes at 42°C. Reactions took place in microtiter dishes so that several could be performed at once using a 2 μl repeating dispenser (Hamilton PB6000): 2 μl of annealed mix per well, 2 μl of dilute AMV RT per well, and 2 μl of dideoxynucleotide mix per well. Reagent concentrations were as published by Hamlyn et al. [15] with some modifications. We did not incorporate isotope during the reactions but instead used kinased oligonucleotide primers. We used higher nucleotide concentrations, but the same ratio of nucleotides to dideoxynucleotides (i.e., [dNTP]:[ddNTP] is 200 μM:40 μM rather than 50 μM:10 μM). Our annealed mix is 0.2 mg/ml in viral RNA or 5 mg/ml in total infected-cell RNA and 0.2 μM in oligonucleotide. Reactions were incubated at 42°C (usually) for 30 minutes, chased with cold nucleotides for 15 minutes, stopped with 4 μl/well formamide dye buffer, heated at 90°C for 5 minutes, and loaded onto thin gradient electrophoresis gels as described [13]. Gels were run at 40W (bringing them to approximately 65°C) for 3 hours.

Additional Denaturants

Various denaturants were used to overcome problems with template and product secondary structure. Viral RNA at 1 mg/ml was treated with 20 mM methyl mercury (Alpha Products, MD) for 10 minutes at room temperature, then treated with 0.1 mM β-mercaptoethanol for 10 minutes at room temperature, then ethanol precipitated, resuspended, and used directly for sequencing. Alternatively, T4 gene 32 protein (Boehringer Mannheim) was added at 1 μg/ml to sequencing reactions and removed with proteinase K as described [4]. In other trials, template RNA was treated with 90% dimethyl sulfoxide, annealed in 18% dimethyl sulfoxide, and the reactions were carried out with a residual 1.8% dimethyl sulfoxide as described [3]. High temperature was used as a denaturant without changing the conditions of our routine sequencing reactions as described above. To overcome problems with product secondary structure, sequencing reactions were run on thin polyacrylamide gels that were similar in buffer composition to the routine gels, but deionized formamide was substituted for water, resulting in 98% formamide [6].

Results

We encountered two regions of problematic secondary structure in sequencing the single-stranded RNA of LCMV. Both were intergenic regions: the first was a 21-base hairpin of $\Delta G = -47$ kcal, and the other was a more difficult structure for which we have no simple free energy calculation. The first secondary structure region was sequenced using the usual reaction conditions and running the products on a denaturing formamide gel as described in Methods. The more difficult structure withstood a variety of approaches. Treatment of the template with methyl mercury did not allow us to obtain sequence past the "knot" of premature terminations. Inclusion of T4 gene 32 protein in the sequencing reaction worked to some degree, that is, a faint but discernible sequence could be detected beyond the secondary structure. This single-stranded binding protein was recommended for DNA sequencing at 37°C [4] and is somewhat useful for RNA sequencing at 42°C as well. Another denaturant, dimethyl sulfoxide, allowed us to read some sequence beyond the secondary structure, but the quality of sequence was similar to that obtained

with T4 gene 32 protein. The clearest sequence information beyond the secondary structure was obtained with high temperature reaction conditions. As can be seen from Figure 1, sequence reactions at 42°C resulted in premature termination by the AMV RT, whereas sequencing reactions at 47°C and 50°C, although they also had premature terminations, extended sufficiently through the region of secondary structure that a readable sequence ladder could be discerned.

Figure 1. Dideoxynucleotide sequence analysis of LCMV RNA at different reaction temperatures: 42°C, 47°C, and 50°C. This intergenic region of the viral RNA was problematic in that it caused premature termination of the AMV RT reactions as well as sequence compressions.

We were concerned that under our reaction conditions, the optimum temperature for nucleotide incorporation by AMV RT was no longer 42°C. We performed incorporation assays at 37, 42, 47, 50, and 55°C and found that optimum incorporation was still 42°C, with half as much incorporation at 50°C and 10% incorporation at 55°C. Nevertheless, we were able to read a crisp sequence from reactions run at 47°C, 50°C (Figure 1), and even 55°C (now shown).

To overcome problems with product secondary structure, nucleotide analogues are frequently used. The incorporation of deoxyinosine-5′-triphosphate (dITP), a guanine analogue, requires that separate ddNTP mixes be made containing a 4:1 mixture of dITP:dGTP in place of the usual dGTP [14]. Similarly, 7-deaza-dGTP (7-deaza-2′-deoxyguanosine-5′-triphosphate) can be substituted for dGTP at a molar ratio of 1:1 (dGTP:7-deaza-dGTP) [7]. The inclusion of formamide in the sequencing gel also denatures product secondary structure and prevents sequence compression [6]. The formamide gels require more time and Temed for polymerization as well as more time to run, but are just as effective as the guanine analogues in solving the problem of product secondary structure. We chose to use formamide gels to solve this problem rather than maintaining separate ddNTP mixes for these purposes, but this is up to the individual investigator.

Discussion

The breakthrough in the use of AMV RT for sequence analysis was the discovery [15, 16] that this enzyme has a greater affinity for the dideoxy analogues than does the DNA polymerase originally used for enzymatic sequencing [17]. The ddNTP mixes for DNA polymerase catalyzed reactions could not be used for AMV RT; instead, the concentration of ddNTP chain terminator in each mix had to be reduced at least 50-fold such that the ratio of dNTP:ddNTP was 5:1 rather than 1:10. Buffer and temperature conditions had already been optimized for the use of AMV RT in the production of cDNA [2, 18], so it was a reasonable advance to use an RNA-dependent polymerase for chain termination sequencing. This method quickly surpassed the chemical sequencing methods in use for RNA sequence anal-

1989, Gene Anal Techn 6:25–28

ysis simply because there were fewer procedures involved. A second advance was the use of end-labeled primers for sequence analysis of impure templates. This meant that template purification steps were eliminated, and sequence could be obtained from total cellular RNA or from chromosomal DNA.

We have used direct RNA sequencing with γ-^{32}P-labeled oligonucleotide primers in a stepwise manner. The first oligonucleotide primer we used depended upon a chemical sequence determination of the 3′ end of the viral genomic RNA [9]. Subsequent primers were synthesized according to sequence obtained from the previous primer extension chain-termination reactions. We encountered two regions of extensive secondary structure, one was a 21-base hairpin in the region between the glycoprotein and nucleoprotein genes of LCMV, and the other was in the region between the L protein gene and a newly discovered region of 500 bases at the 5′ end of the genome (M. Salvato and E. Shimomaye, in preparation). The first region (a 21-base intergenic hairpin) has a calculated free energy of −47 kcal/mole [19] and was sequenced at the usual temperature (42°C) but the reactions were run on 98% formamide gels. The second region of intergenic structure proved to be much more difficult. Although it has a calculated free energy of only −30 kcal, it is probably involved in tertiary base pairing such as the pseudoknot structures predicted for viral RNAs [20, 21], and sequence could only be obtained by the use of denaturant chemicals or by high reaction temperatures, it was necessary to use 98% formamide gels to avoid the sequence compressions caused by intrachain base stacking and pairing.

The observation that AMV RT terminates within intergenic regions at a physiologic temperature suggests an involvement of RNA structure in the regulation of viral transcription. It is possible that intracellular denaturants, such as the T4 single-strand binding protein, enable transcription to proceed through regions of RNA higher order structures during viral replication but are nonfunctional during messenger transcription.

This is publication number 5423-IMM from the Department of Immunology, Scripps Clinic and Research Foundation, La Jolla, California. This work was supported in part by USPHS grants AI-09484 (MS and M.B.A. Oldstone) and AI-25522 (MS).

References

1. Bailey, J. M., and Davidson, N. (1976) Anal. Biochem. 70, 75–85.
2. Buell, G. N., Wickens, M. P., Payvar, F., and Schimke, R. T. (1978) J. Biol. Chem. 253, 2471–2482.
3. Bassel-Duby, R., Spriggs, D. R., Tyler, K. L., and Fields, B. N. (1986) J. Virol. 60, 64–67.
4. O'Farrell, P. (1987) Boehringer Mannheim Biochemica 4, 13.
5. Huibregste, J. M., and Engelke, D. R. (1986) Gene 44, 151–158.
6. Murchie, M. J., and McGeoch, D. J. (1982) J. Gen. Virol. 62, 1–15.
7. Mizusawa, S., Nishimura, S., and Seela, F. (1986) Nucl. Acids Res. 14, 1319–1324.
8. Tabor, S., and Richardson, C. C. (1987) Proc. Natl. Acad. Sci. USA 84, 4767–4771.
9. Auperin, D. D., Compans, R. W., and Bishop, D. H. L. (1982) Virology 121, 200–203.
10. Peattie, D. (1979) Proc. Natl. Acad. Sci. USA 76, 1760–1764.
11. Maxam, A. M., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.
12. Maniatis, T., Jeffrey, A., and van de Sande, H. C. (1975) Biochemistry 14, 3787–3794.
13. Biggin, M. D., Gibson, T. J., and Hong, G. F. (1983) Proc. Natl. Acad. Sci. USA 80, 3963–3965.
14. Bankier, A. T., and Barrell, B. G. (1983) in Techniques in Life Sciences, B5: Nucleic Acids Biochemistry (Flavell, R. A., ed.), pp. 1–34, Elsevier, Ireland.
15. Hamlyn, P. H., Brownlee, G. G., Cheng, C.-C., Gait, M. J., and Milstein, C. (1978) Cell 15, 1067–1075.
16. Both, G. W., and Air, G. M. (1979) Eur. J. Biochem. 96, 363–372.
17. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
18. Wickens, M. P., Buell, G. N., and Schimke, R. T. (1978) J. Biol. Chem. 253, 2483–2495.
19. Tinoco, I., Borer, P. N., Dengler, B., Levine, M., Uhlenbeck, O., Crothers, D., and Gralla, J. (1973) Nat. New Biol. 246, 40–41.
20. Pleij, C. W. A., Reitveld, K., and Bosch, L. (1985) Nucl. Acids Res. 13, 1717–1731.
21. Clarke, B. E., Brown, A. L., Curry, K. M., Newton, S. E., Rowlands, D. J., and Carroll, A. R. (1987) Nucl. Acids Res. 15, 7067–7079.